DOCUMENT RESUME

ED 460 954 TM 028 897

AUTHOR Neill, Monty
TITLE Testing Our Children: A Report Card on State Assessment Systems. Executive Summary and Summary of State Findings.
INSTITUTION National Center for Fair and Open Testing (FairTest), Cambridge, MA.
SPONS AGENCY Joyce Foundation, Chicago, IL.; Ford Foundation, New York, NY.
PUB DATE 1997-09-00
NOTE 43p.; FairTest Staff contributors were Laura Barrett, Phyllis Bursh, Mark Earley, Charles Rooney, Robert Schaeffer, Leslye Sneider, and Marilyn Yohe. For the full report, see TM 028 896.
AVAILABLE FROM National Center for Fair and Open Testing (FairTest), 342 Broadway, Cambridge, MA 02139 ($10). Tel: 617-864-4810; Fax: 617-497-2224; Web site: http://www.FairTest.org.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; Decision Making; *Educational Assessment; Educational Change; Elementary Secondary Education; Objective Tests; Report Cards; Standardized Tests; *Standards; *State Programs; Surveys; Test Bias; *Test Use; Testing Problems; *Testing Programs
IDENTIFIERS *FairTest

ABSTRACT

FairTest evaluated how well state assessment practices live up to the promise of high standards without standardization. The practices of states were measured against standards derived from the "Principles and Indicators for Student Assessment Systems," a 1995 publication of education and civil rights groups working through the National Forum on Assessment. FairTest used surveys, interviews, and various documents to evaluate the states and developed a scoring guide to evaluate each state. Survey responses were received from 44 states, and FairTest drew on other documents to evaluate the other 6 states. It found that, after nearly a decade of intensive discussions about the role and nature of assessment, and despite some important improvements, the fundamental approach of state testing has not changed. Labels have sometimes been revised to "assessment," but most state programs still rely on traditional, multiple-choice tests, and most states still use them inappropriately to make high-stakes decisions. Two-thirds of state student assessment systems do not even reach the middle level of system quality. One-third of systems need a complete overhaul, and another third need major improvements. In two-thirds of the states it may be said that testing systems often impede, rather than enhance, genuine education reform. Many states do not base their assessments on their content standards, and too many states use norm-referenced tests rather than tests that compare achievement to state standards. State findings are summarized, the standards and scoring guide are discussed, and a state data table is presented. (SLD)

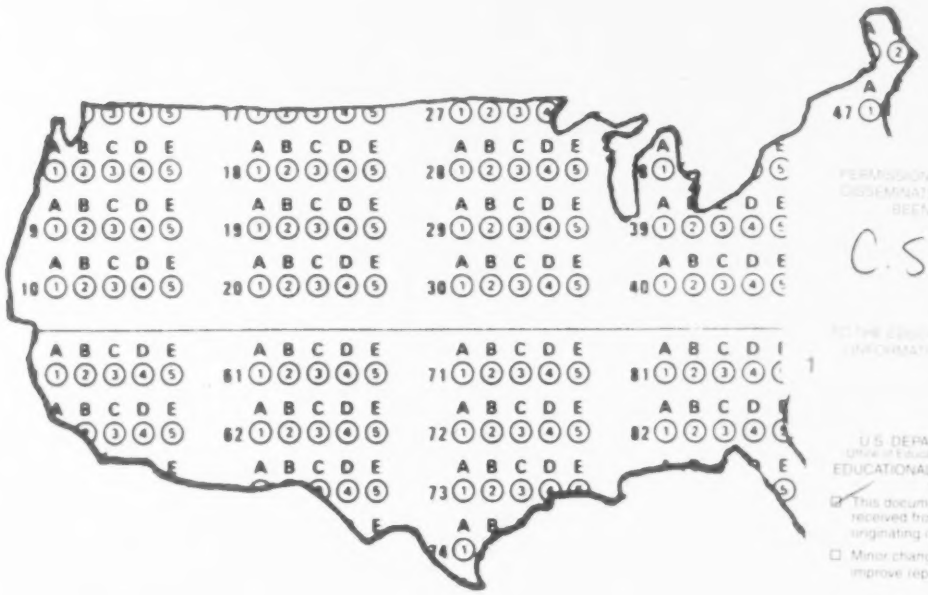
Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 460 954

TESTING OUR CHILDREN

A Report Card on State Assessment Systems

Executive Summary and Summary of State Findings



PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

C. Schuman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM028897

by Monty Neill and the Staff of FairTest

FairTest: The National Center for Fair & Open Testing

BEST COPY AVAILABLE

Executive Summary and Summary of State Findings

TESTING OUR CHILDREN

A Report Card on State Assessment Systems

by Monty Neill, Ed.D.

FairTest Staff who contributed to this report:
Laura Barrett, Phyllis Bursh, Mark Earley, Charles Rooney, Robert Schaeffer, Leslye Sneider, Marilyn Yohe

September 1997
$10.00

For a copy of the full report, including all state evaluations, use the order form at the end of this report.

National Center for Fair & Open Testing (FairTest)

342 Broadway
Cambridge, MA 02139
tel: 617/864-4810 fax: 617/497-2224
email: FairTest@aol.com
web page: www.FairTest.org

Testing Our Children: A Report Card on State Assessment Systems
© 1997 National Center for Fair & Open Testing

This report was made possible by grants from the Joyce Foundation and the Ford Foundation. The opinions and conclusions expressed in the report are solely those of FairTest.

Thanks to Steve Ferrara, Ed Reidy and Ramsay Selden, who offered comments at various points. They are not responsible for the conclusions or any errors that may be found in this report. Thanks also to the staff of the North Central Regional Education Laboratory and the Council of Chief State School Officers who facilitated our receipt of their data.

Testing Our Children:

A Report Card on State Assessment Systems

Executive Summary and Summary of State Findings

Table of Contents

Introduction 1

Executive Summary 5

A. Findings in Brief 5
B. State Performance Levels 7
C. Patterns and Trends 8
D. Recommendations 10

State Findings 13

A. Summary of State Findings 13
B. Standards 30
C. Scoring Guide 33
D. State Data Table 34

Methodology 39

Testing Our Children:
A Report Card on State Assessment Systems

Introduction

Standardized tests first rose to prominence in the 1920s, the era in which the "factory model" of education established clear dominance. They reinforced that mode of schooling, in which only a few children received a high-quality education, and they were used to sort students hierarchically within that model. The promise of school reform in the 1990s has been to break with that inadequate, often harmful model of schooling. As one part of reaching that goal, assessment must be fundamentally restructured to support high standards without standardization.

In this study, FairTest evaluates how well state assessment practices live up to this promise. We have measured these practices against standards derived from the *Principles and Indicators for Student Assessment Systems,* a 1995 publication by a coalition of education and civil rights groups working together through the National Forum on Assessment.

In broad terms, the *Principles* calls for assessments that are:
- grounded in solid knowledge of how students learn;
- connected to clear statements of what is important for students to learn;
- flexible enough to meet the needs of a diverse student body; and
- able to provide students with the opportunity to actively produce work and demonstrate their learning.

What we have found is that despite nearly a decade of intensive discussions about the role and nature of assessment, and despite some important improvements, the fundamental approach of state testing programs has not changed. Though the labels have often been revised to "assessment," most state programs still predominantly rely on traditional, multiple-choice tests, and many states use them inappropriately to make high-stakes decisions.

Based on a detailed survey and other data sources, we conclude that two-thirds of state K-12 student assessment systems do not reach even the middle level of system quality. One-third of the systems need a complete overhaul, and another third need major improvements if they are to provide support for high quality teaching and learning. The remaining third all have positive components, but still need some improvements.

In two-thirds of the states, then, testing systems often impede, rather than enhance, genuine education reform:

- Rather than holding schools accountable for providing a rich, deep education and reporting on such achievement to the public, most state testing programs provide information on a too-limited range of student learning in each important subject area.

• Rather than supporting and assessing complex and critical thinking and the ability to use knowledge in real-world situations, most state tests continue to focus too much on measuring rote learning.

• Rather than making decisions about students based on multiple sources of evidence, too many states use a single test as a mandatory hurdle.

Since state tests powerfully affect curriculum and instruction, most state testing programs present obstacles to developing high-quality classroom practices and fail to support strong school reform. Some improvements can be seen in the use of writing samples (though these are often themselves narrow) and constructed-response items (though their use remains too limited), and in more attention to bias reduction. However, in most states, these modest changes amount to tinkering at the edges of reform.

In fact, the recent tendency has been to intensify the traditional mode of testing, with higher cut-off scores and more "difficult" exams, without changing the underlying approach. In most state tests, "difficult" means testing student achievement in conventional academic subjects at an earlier age, such as algebra in grade 8. The problem with this approach is not that algebra now may need to be taught in grade 8, but that the kind of algebra tested remains predominantly the memorization of rules and procedures and very limited applications. This approach fails to meet the essence of the math standards of the National Council of Teachers of Mathematics. A similar, flawed approach can be found for every subject.

The negative consequences of relying on traditional tests and using them to control school reform often seem to be the result of continued confusion over the limitations of large-scale assessments. Unfortunately, states often fail to recognize these limitations and expect their tests to be useful in ways they cannot.

Large-scale testing programs are generally not useful in improving a student's immediate learning process, though clearly that is what most parents hope for from assessment. As diagnostic tools, most large-scale tests are blunt, imprecise, and often useless -- but most states claim that diagnosis is a reason for their tests. Because most state tests do not provide any opportunity for sustained and engaged thinking, they are poor tools for shaping or improving curriculum and instruction -- a goal most states claim for their tests. While these exams can provide some information to the public about what students have learned, most do not provide information about whether students can use in their lives the things they have supposedly learned. They thus provide limited accountability information.

Despite these extreme limitations of state testing programs, the cumulative effect of the multiple uses of these tests is that the exams largely define the purpose and processes of schooling in most states. They affect not only curriculum and instruction, but also the culture of learning, student motivation, and the underlying conceptions of what learning is and how humans learn. Driving school reform with traditional tests will not succeed if the nation really wants all children, not just the children of the wealthy, to gain an education that challenges

their minds and spirits, that assumes not only that they can learn some skills but can learn to use their learning as active participants in a democratic society.

There is an alternative. The *Principles and Indicators* calls for large-scale assessments that combine sampling from classroom-based assessment data, such as portfolios and learning records, with performance exams administered to samples of students. In this way, essential standards are promoted and accountability information is gathered, while schools are encouraged to become communities of learning that support all their students. Only one state, Vermont, approaches this model, though elements of the assessments in a few other states are headed in this direction.

Fundamental assessment reform is still feasible. What is lacking is not the technical know-how, though much remains to be learned in that domain, but the political will. The responsibility for improving assessment programs rests first of all with policymakers -- governors, legislators, boards of education. It rests secondly with all those who can educate, or influence, the policymakers -- educators, parents, community and business leaders, testing experts, state education staff, and the voting public. That makes achieving real assessment reform an education and organizing project. Only with an informed and active community, as well as educated policymakers, can deep reform be created and sustained, including the necessary transformation of state assessment programs.

Executive Summary:
State assessment systems in light of the *Principles and Indicators for Student Assessment Systems*

Across the nation, state testing systems powerfully affect curriculum, instruction, school cultures, and the quality of education delivered to our nation's children. They can either support important learning or undermine it.

This study evaluates how well state assessment systems support and help improve student learning. FairTest based its evaluation on standards derived from the *Principles and Indicators for Student Assessment Systems.* This document was developed by the National Forum on Assessment to help guide assessment reform and has been signed by over 80 education and civil rights groups. To gather data, FairTest used surveys, follow-up interviews, and various documents.

A. Findings in Brief

Among the findings of this study are the following:

1) On a five-point scale for scoring state assessment systems, two-thirds of state K-12 student assessment systems do not reach even the middle level of system quality: one-third of the systems need a complete overhaul and another third need major improvements if they are to provide support for high quality teaching and learning. A few states have made good progress, reaching level 4, but only one, Vermont, has reached the top level.

2) While most states now have content standards, many state tests are not based on their standards, and many important areas in their standards are not assessed.

3) Most states rely far too heavily on multiple-choice testing and fail to provide an adequate range of methods for students to demonstrate their learning. This results in not assessing important areas and creating the likelihood that those areas will not be taught.

4) Too many states use norm-referenced tests (NRTs), which compare students to reference groups and not to achievement on state standards. These tests fail to assess important areas of the standards and encourage grouping and instructional practices that historically have failed to provide many students with a strong education.

5) The state testing burden is often too heavy, with students repeatedly tested in the same subjects. A few states test students in almost every grade. For accountability purposes, extensive testing is not necessary.

6) Seventeen states use a single test as a necessary requirement for high school graduation, violating the AERA/APA/NCME standards for good assessment practice, ensuring unfair treatment of many students, and increasing the likelihood that narrow tests will dictate curriculum and instruction. Districts may use state tests as graduation or grade promotion hurdles. An additional five states currently plan to implement such tests, two of which plan to allow an alternative option.

7) Most writing assessments require students to respond to a single prompt, fostering and reporting a limited conception of writing. Writing must serve many purposes and therefore take many styles. A major problem here is the potential reduction of writing instruction to fit the state exam.

8) Rich assessment techniques, such as portfolios and performance events, are rarely used by states. Thus, important areas of learning are not assessed and important signals are not sent to schools about what students should be learning and how assessment can support that learning.

9) Very few states use sampling for accountability, public reporting, and program improvement purposes, even though it provides accurate data, is less expensive and less intrusive, and allows greater use of portfolios and performance events.

10) Most states use tests for student diagnosis and for improving curriculum and instruction, even though most large-scale tests are crude tools for diagnosis and too narrow to support high quality curriculum and instruction.

11) A solid majority of states have bias review panels, often with significant authority to delete or revise items on state-made tests, but some do not. This is a positive development.

12) States tend not to adequately assess or include in state reports students with Individual Education Plans (IEP, e.g., "special education") and students with Limited English Proficiency (LEP). Inclusion of all categories of students, using appropriate assessments, is necessary for proper program evaluation and ensuring proper education for these students. The recently reauthorized federal Individuals with Disabilities Education Act will require all students with disabilities to be assessed appropriately, but such provisions do not exist for LEP students.

13) States are generally quite weak in providing adequate professional development in all aspects of assessment to teachers and other educators. Such teacher education, particularly in classroom assessment, is fundamental to assessment and broader school reform.

14) Few states evaluate teacher competence in assessment or study district, school and classroom assessment practices or their impacts. Thus, they lack information to help improve the quality of assessment at all levels and to halt harmful practices.

15) Student and parent rights, such as the ability to review tests after completion, to challenge flawed items or to appeal scores, exist unevenly. Such rights are fair in themselves and also

help parents better understand assessment and education in general and to view themselves as important partners in their children's education.

16) Reporting to the public and educating the public about assessment are often limited, and few states report in languages other than English, even if they have a large number of residents who do not speak or read English.

17) State reviews of their assessment systems need substantial improvement. Most do not study the impact of testing on curriculum, instruction, or graduation rates; and most do not review whether their assessments measure the ability of students to think critically or in complex ways in the various subject areas. In an era in which testing is proposed as a fundamental tool for school reform, states often cannot even be sure whether increasing scores are based on real learning gains or teaching to the test.

B. State Performance Levels

Using a scoring guide, FairTest evaluated each state. The list below reports which states scored at each level of the scoring guide. The scoring guide is found in the section on state findings, and details for each state are provided in the full report.

How the States Scored	
Level 5	**A model system.** Vermont
Level 4	**State assessment system needs modest improvement.** Colorado, Connecticut, Kentucky, Maine, Missouri, New Hampshire
Level 3	**State assessment system needs some significant improvements.** Illinois, Kansas, Maryland, Michigan, Oregon, Pennsylvania, Rhode Island
Level 2	**State assessment system needs many major improvements.** Arkansas, California, Idaho, Indiana, Massachusetts, Minnesota, Montana, Nebraska, Nevada, North Dakota, New Jersey, New York, Ohio, Oklahoma, South Dakota, Texas, Washington, Wisconsin
Level 1	**State assessment system needs a complete overhaul.** Alabama, Alaska, Arizona, Florida, Georgia, Hawaii, Louisiana, Mississippi, New Mexico, North Carolina, South Carolina, Tennessee, Utah, Virginia, West Virginia
	Not scorable. Delaware, Iowa, Wyoming

BEST COPY AVAILABLE

C. Patterns and Trends

A few basic patterns and trends over the past decade, based on a comparison between this and other reports, can be discerned. These include:

1) The amount of testing done by the states appears not to have changed very much, though it seems to vary year to year as states alter their testing programs.

In its 1988 report, *Fallout from the Testing Explosion*, FairTest found, by comparing the numbers of tests administered to school enrollments, that states were administering .42 tests (which may include more than one subject area) per year per student. (District testing, primarily achievement and special needs testing, raised the average to about 2.5 tests per student per year.)

To identify current testing frequency, we examined CCSSO/NCREL data for various grades tested over the past few years. The 1993-94 data show that the states tested a total of 278 grades, or an average of 5.56 grades. (This assumes a state uses only one test at a grade level, but some do use more than one test at a given grade level). With 13 grades, this averages to .43 tests per year per student. In 1994-95, the numbers declined to 243 grades tested, or an average of 4.86 grades or .37 tests per year per student. But in 1995-96, the numbers were back up slightly, to 264 grades tested, or 5.28 tested grades per state and .41 tests per student.

As the means of determining the amount of testing was different in *Fallout*, the numbers are not directly comparable, but they give a rough sense of the stability of the amount of state testing over time.

2) *Fallout* reported that 11 southern states (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee and Virginia) tested more often than did the rest of the nation. This continues to be true. In 1995-96, those 11 states tested in 7 grades on average. The other states which are part of the Southern Regional Education Board (SREB) actually now test even more: Texas tested at 9 grades, Maryland at 8, Oklahoma at 8, and West Virginia at 11, bringing the SREB average to 7.5 grades, substantially higher than the national average of 5.28 grades. Another way of looking at it is that 30 percent of the states do 43 percent of the testing.

3) These states are also more likely to mandate high school graduation tests. Of the 15 SREB states, 11 have graduation exams. Only six of the 35 states outside the South have such a test.

4) The number of states with high school exit exams declined in the 1990s but is now growing again. In 1989, *Education Week* (May 10) reported 23 states had or intended to have these exams. By 1994-95, CCSSO/NCREL reported that 17 states had mandatory exit exams. FairTest confirmed this number, but also found that five more states plan to adopt such a requirement.

5) Other than southern states, half the states with high school exit exams are in the northeast: New Jersey, New York and Ohio are joined by Hawaii, New Mexico and Nevada. The states that soon will require such tests are Alaska, Arkansas, Delaware, Indiana, and Massachusetts. This will bring the total number of states that have or are planning to have exit exams to 22 -- about where it was at the end of the 1980s.

6) *Fallout* noted that large cities tested more often than smaller cities or rural areas. Combined with the data on southern states, this suggests that areas with large proportions of African Americans are most likely to test heavily. States with relatively large proportions of African Americans are more likely to administer high school exit exams.

7) It also appears that the 15 SREB states, with the notable exception of Kentucky and Maryland, are less likely to use constructed-response or performance assessments (excepting writing to a prompt) than is the nation as a whole. States with mandatory high school exit tests also appear less likely to use constructed-response or performance assessments, again excepting writing to a prompt (see Fairbanks & Roney). These findings may be starting to change as more states use constructed-response items, including in graduation tests.

8) Southern states also are more likely to use NRTs. Thirty-three states use an NRT, including those which sample (North Carolina and Maryland), those which require it of districts (Nebraska) or pay for districts use of one (California and Iowa). All of the 15 SREB states except Texas use an NRT. Roughly half of the remaining states use an NRT (19 of 35).

9) All told, there appears to be a "southern effect" which includes high-stakes testing, a heavy testing load, use of an NRT, and relatively less use of constructed-response and performance assessments. As a group, the southern states still are the nation's poorest region, so this is also a "poverty effect." Results of the National Assessment of Educational Progress continue to show the southern region lagging behind the rest of the nation in terms of measured educational achievement.

Since there is evidence that using performance assessments signals or spurs a shift toward teaching and assessing more challenging, cognitively complex material, then the southern states could be left behind once again. As the negative effects of teaching to narrow tests most powerfully affect schools with large proportions of minority-group and low-income children, such students in these states are particularly at risk of continuing to receive a low-level education that will not prepare them well for their adult lives. Students in large cities that also emphasize teaching to traditional tests face the same risk.

Unfortunately, these southern states, along with others, are caught in a vicious circle. Low scores lead to more tests and higher stakes. More tests and higher stakes lead to more intense "teaching to the test." Teaching to narrow, multiple-choice tests leads to an overemphasis on rote memorization at the expense of higher order thinking skills. In this way, tests themselves are part of the problem, not the solution.

Fortunately, several states across the country are trying to break this cycle. They are increasing their use of assessments that measure genuine knowledge, not simply facts, and that evaluate a student's performance on multi-faceted tasks, not simply his or her ability to select the preferred response from a list of possible answers. They are also paying great attention to professional development so that teachers learn well how to use performance assessments and portfolios in their classrooms. This facilitates a bottom-up approach to school reform rather than relying solely on top-down, test-driven initiatives.

If these alternative assessment systems are allowed to survive the growing pains of their early years, they will provide educators in other states with valuable knowledge about how to alter their assessment systems. Perhaps then most of the states, not just a few, will move beyond tinkering at the margins and will completely overhaul their state assessment systems.

D. Recommendations

These findings establish the framework in which fundamental assessment reform must take place. A great deal has been learned, some of it from pioneering efforts in a few states, some of it in districts, most of it in schools and classrooms. What is lacking is not the technical know-how, though certainly problems remain, but the political and social will to recreate assessment as part of reinventing education.

If large-scale assessments are to support excellence and equity in education, FairTest concludes that underlying conceptions and basic practice in most states need to be fundamentally changed and brought into alignment with the *Principles and Indicators for Student Assessment Systems* as follows:

1) Base all state (or district) assessments of student achievement on clear standards.

2) Employ multiple methods of assessment, limiting multiple-choice to no more than one quarter of test-takers' scores.

3) Rely on methods that allow students to demonstrate understanding by applying knowledge and constructing responses and that ensure assessment of complex and critical thinking in and across subject areas.

4) Do not use norm-referenced tests, or limit their use to very light sampling.

5) Do not make high-stakes decisions, such as high school graduation, using single exams as a hurdle. Rely on multiple sources of information.

6) Employ sampling procedures to collect information on large populations, using performance and portfolio assessments.

7) Rely on sampling from classroom-based work as a key component of large-scale information on student achievement, including work which allows individual choices and expressions of knowledge and provides students the opportunity to evaluate their own work.

8) Enhance efforts to appropriately include all students in assessments and reporting, and report disaggregated data by important population groups.

9) Ensure adequate professional development in assessment, particularly in classroom and performance assessment, for both teachers and students in education schools.

10) Systematically involve teachers and other educators in developing and scoring performance assessments and portfolios.

11) Institute comprehensive reviews and use the results to improve assessments.

State Findings

To evaluate the specific characteristics of state assessment programs, FairTest adapted the *Principles and Indicators* to create standards and indicators appropriate for large-scale assessment. The standards are:

Standard 1: Assessment supports important student learning.
Standard 2: Assessments are fair.
Standard 3: Professional development.
Standard 4: Public education, reporting, and parents' rights.
Standard 5: System review and improvement.

The following explains the basic purpose of each standard and indicator and why it is important, summarizes the findings from across the states, and discusses the implications of each finding. Forty-four states responded to the FairTest survey, providing relatively complete information for the evaluation process. For the remaining six states, FairTest relied on other sources which provided substantially less data and no information at all on many of the indicators in the standards.

A. Summary of State Findings

Standard 1: Assessment supports important student learning.

The *Principles* states: "Assessment systems provide useful information about whether students have reached important learning goals....They employ practices and methods that are consistent with learning goals, curriculum, instruction, and current knowledge of how students learn. No assessment... is used that narrows or distorts the curriculum or instructional practice."

Large-scale assessments should be used to gather data for program improvement and to report program-level data to the public. Most other assessment purposes, such as individual student diagnosis, reporting individual progress and determining who should graduate, are better left to schools and teachers. Large-scale assessments are necessarily blunt instruments, and so should be used sparingly, with caution, and for purposes in which large-scale information makes sense.

Unfortunately, state programs often undermine important student learning through overuse of multiple-choice testing and norm-referenced tests, under-utilization of performance assessments and portfolios, high-stakes uses of single exams, and over-testing. The assessments are often so limited as to undermine content standards (which most states have adopted) by not assessing important areas in the standards. Though one of the most commonly stated purposes of state assessments is "program improvement," most state assessments are not adequate for helping to develop high-quality education programs.

Some states do not have state testing programs. The *Principles* does not recommend either state standards or state assessments and recognizes these can be undertaken at the district level. However, FairTest concludes that states which rely on district testing should then evaluate district practices and support improvements at the district level. In some states without formal state programs, the state mandates district assessments. In these cases, the mandate is effectively a state program and can be evaluated as such. The state also can be evaluated in terms of its direct activities or support for districts on the issues of fairness, professional development, reporting, and evaluation of the assessment program.

1.1. Assessments are based on and aligned with standards. Students deserve to have clear statements of what they are expected to learn and the opportunity to master that material. States should have standards if they have state exams, and the exams should assess comprehensively and in a balanced fashion the content that is in their standards. If a state mandates district achievement testing, it also should mandate that those tests be based on state or district standards.

While most states now have standards and increasingly report that their assessments are aligned to the standards, too often important areas in these standards are not assessed. This is largely because of limited assessment methods, particularly over-reliance on multiple-choice testing. Some states acknowledged this, noting such things as "multiple-choice cannot assess all areas in the standards" or even noting a percentage of the standards that is measured. Others simply claim that their multiple-choice tests are matched to the standards. The reality is that most state tests do not comprehensively and in a balanced manner assess students to high-quality content standards.

The clear dangers are that what is not tested is not taught, that what is tested is the lower levels of the standards, and that curriculum is therefore reduced to its lower levels. Based on previous experience, the curriculum is most likely to be narrowed in schools and districts where students do not perform as well on the tests. The consequence, which has been observed in various research studies, is often to continue to deny a challenging and engaging education to those students who have historically not been well-served by public schooling, particularly students from low-income families and students of color. As discussed in Standard 5, it appears that few states seriously investigate this issue.

1.2. Multiple-choice and very-short-answer (e.g., "gridded-in") items are a limited part of the assessments; and assessments employ multiple methods, including those that allow students to demonstrate understanding by applying knowledge and constructing responses. These requirements are strongly stated in the *Principles*. FairTest recommends that not more than one quarter of a student's score in any subject be obtained from multiple-choice and very-short-answer items.

Serious critical and complex thinking in subjects, real-world problem solving, and application of knowledge cannot be assessed adequately with multiple-choice items. Further, as teachers tend to teach to state exams, focusing instruction on multiple-choice tests limits

curriculum and instruction in ways that deny students opportunities to think, tends to narrow the range of instructional practices, and reduces student motivation to learn -- all of which combine to undermine both excellence and equity. Using such tests for "diagnosis," as many states report doing, compounds the problem: they are too limited a measure for useful diagnosis for most instructional purposes.

Most of a score should come from methods that allow students to apply knowledge, solve complex problems, and demonstrate thinking within a subject. Such an approach enables assessment to better match high-quality standards. These are also practices that are more compatible with how humans learn. Additionally, using multiple methods allows students with different learning styles an opportunity to demonstrate their achievement and enables the assessment of content or skills that are not assessed well by other methods.

Unfortunately, most states rely too heavily on multiple-choice items and fail to use a reasonable range of assessment methods. Excluding writing assessments, of the 50 states, 26 rely entirely or nearly entirely on multiple-choice. Another 16-18 rely mostly on multiple-choice (have less than half their scores derived from constructed-response items; in two states, the proportions were not clear but appear to be around the one-half point). Only 6-8 states have less than half multiple-choice items.

Using a variety of methods does not require that multiple-choice be one of them. Rather, the mix could include short and extended constructed-response items, performance events, and portfolios.

Most fundamental is that the actual tasks and items are of high-quality. This study could not evaluate the quality of the items or whether taken together they comprise a high-quality assessment.

Thirty-eight states have writing assessments (including Vermont, where it becomes mandatory next year). However, with rare exceptions, the writing is simply responding to a pre-selected prompt, with students allowed no opportunity even to select from a set of prompts. Only three have portfolio writing assessments. Unfortunately, response to a prompt creates a very narrow picture of writing and encourages teaching geared to an arbitrary formula, such as the five-paragraph "essay." This is also an equity issue, as students who happen to be interested in or knowledgeable about the one particular topic will have an unfair advantage. Instead, more than one form of writing should be assessed and students should have a choice of prompts. An additional issue is the time allowed for response, which in some states is too short. Some research suggests that student performance improves with extended time for response, a point that is relevant not just to writing.

1.3. Assessments designed to rank order, such as norm-referenced tests (NRT), are not used or are not a significant part of the assessment system. These tests are constructed to compare students rather than to see how well students achieve according to standards. Norm-referencing is rooted in the concept of the "bell curve." The use of comparisons and the bell

curve, which by definition place half the students "below average" or even "below grade level," suggests that many students will not learn to high levels and meet state standards. The use of NRTs often encourages tracking, sorting and low expectations.

Thirty-three states use NRTs, some as the major state component and some together with a criterion-referenced test (CRT); two of the 33 use them only on a sampling basis. Some NRTs now include, as an option, constructed-response items, but almost all states which use commercial NRTs still use exclusively multiple-choice versions. A few states report their results according to state norms. This is also inappropriate; their exams should be constructed around state standards and be reported in terms of those standards.

1.4. The test burden is not too heavy in any one grade or across the system. Students often are tested far more frequently than is needed to produce data for program improvement or accountability. Consequently, valuable classroom time is wasted preparing for and taking exams that serve no useful purpose. A reasonable system is one in which students are assessed in a subject once at each level (elementary, middle, high), as is now required by the federal Title I program. A model system would rely on sampling.

The test burden required by states varies greatly, from a few tests in a few grades, to many subjects tested in a few grades, to a few subjects tested in many grades, to many subjects tested in many grades. The state test burden is often unnecessarily heavy. Many districts add yet more standardized tests to the state exams, so what appears to be a reasonable burden in some states may be, in most of that state's districts, a high burden. Few states, however, even survey district assessment practices.

FairTest has not addressed the issue of how many subjects should be tested but recommends that if more than two subjects are tested, the burden should be spread over several grades (e.g., English language arts and math in grade 4, science and social studies/history in grade 5). Except for comments in a few state reports, we also did not address the issue of the amount of time devoted to testing.

1.5. High-stakes decisions, such as high school graduation for students or probation for schools, are not made on the basis of any single assessment. The AERA/APA/NCME *Standards for Educational and Psychological Testing* state at Standard 8.12: "[A] decision or characterization that will have a major impact on a test taker should not automatically be made on the basis of a single test score." Similar statements can be found in numerous other test use guidelines, including the *Principles*. FairTest concludes that no single test should act as a barrier to graduation.

By "single assessment" we mean "hurdle" -- as in a track race in which each and every one must be cleared. Thus, using a test as a stand-alone hurdle means it must be passed for graduation or promotion -- even if there are, as is typical, multiple opportunities to clear the hurdle.

However, 17 states use a test as a high school graduation requirement. Two states include state assessments as part of determining grade promotion. Some districts also may use state assessments in determining grade promotion or graduation, though the information on this is largely anecdotal. States sometimes report the tests are also used for placement purposes, which would include tracking and which certainly can be high-stakes uses. States need to monitor districts to ensure tests are not misused in making decisions.

The number of states with graduation exams has been fairly stable at about 17 for a few years. At the turn of decade, FairTest compiled a list of 24 states that had or intended to have such requirements, so by the middle of the decade substantial progress had been made. However, in the past several years, a stronger push has come from a number of quarters to implement graduation exam requirements. It now appears that by about 2000, at least five more states will have such policies in place.

For students, this is substantially a fairness issue. Individuals should be judged on the basis of their accumulated work, not their score on a one-shot test. Similarly a range of information should be considered in evaluating programs. Decisions should not be triggered solely by results on tests. In fact, for most states which have established potentially serious consequences for schools or districts, such as probation or takeover, scores are one of a number of factors which trigger investigations prior to actions, which is as it should be. At a minimum, states with high-stakes tests for individuals should apply this approach.

A second reason for this standard is that the higher the stakes, the more likely the tests will control curriculum and instruction. Graduation tests are usually entirely or almost entirely multiple-choice, sometimes with a writing sample added in, so the issues raised around multiple-choice tests pertain with most force to these high-stakes exams. Any stakes, starting with public reporting and increasing through a variety of sanctions and rewards for schools or students attached wholly or in part to test results, can begin to cause instruction to focus on the content and method of the tests. If this approach to focusing instruction is to be valid, then the exams must adequately assess the range of knowledge, understanding, skills and abilities that schools seek to teach. In addition, the tests should change every year to prevent narrow teaching to one set of items. Few state exams meet these requirements.

1.6. Sampling is employed to gather program information. Sampling, rather than testing every student with an entire exam, is a reasonable solution to a fundamental quandary in large-scale assessment: how to use time-consuming and expensive performance events and portfolios as a major source of data, given limited funds. Matrix sampling, in which an assessment is divided into parts and each test-taker is administered only one of the parts, can be particularly efficient for exams.

Only a few states make even limited use of sampling. Missouri is probably dropping sampling from its new system, Maine uses sampling in some subjects but may be switching to testing every student, and North Carolina and Maryland use sampling with an NRT. The best case is Vermont, which re-scores samples of student portfolios (in which every student

has a portfolio) to obtain state-level data. However, because it has many small schools, Vermont will not use sampling in its new performance exams, but will test every student.

The essential problem, however, is political -- the perception that parents and the public want every child tested and scored. So long as this remains the policy imperative, it is unlikely that much progress will be made in using instructionally appropriate assessment methods. That is, choosing to test every child inexpensively requires the use of narrow testing methods. This educational cost is generally not explained to the public so as to create an informed discussion of the trade-offs.

The educationally superior alternative is to use large-scale assessments employing statistically sound samples to report program data and to have individual data gathered and reported by schools. Schools also would make high-stakes decisions and certify student achievement, such as for high school graduation.

1.7. The evaluation of work done over time, e.g., portfolios, is a major component of accountability and public reporting data. As emphasized in the *Principles*, students should be evaluated primarily on the basis of their regular classroom work, accumulated over time, rather than on the basis of one-time tests. This enables examination of much richer information than can be obtained from "snap-shot" tests. It also supports fairness by allowing and encouraging a greater variety of student work.

Only six states use portfolios at all as part of the state testing program, though a number of other states are supporting districts and schools in developing portfolios. One obstacle has been the complexity of gathering an appropriate selection of a student's work and evaluating it reliably. The education of scorers to respect diversity while insisting on quality also is essential. Nonetheless, the major obstacle appears to be the political decision that the state should assess each individual student, rather than to sample, thus making use of portfolio assessment for program evaluation and accountability very expensive.

1.8. Students are provided an opportunity to comment on or evaluate the instruction they receive and their own learning. Principle 1 notes that self-reflection is an important element of assessment and learning and should be part of the assessment system. While this is primarily a classroom issue, it has a place in large-scale assessments, for two reasons. First, its inclusion signals that self-reflection is important. Second, the information received can be used in evaluating what works and why in curriculum and instruction.

Only a few states include this option, usually in a survey attached to the state exam. Similarly, only a few states survey teachers or administrators about instruction and assessment (see Standard 3).

1.9. Appropriate contextual information is gathered and reported with assessment data. Such data includes information about the actual curriculum and instruction provided to students, the instructional and physical resources, demographic data, information on spending

and the teaching force, class size, student mobility, tracking and placement policies, and other outcome information.

It appears that few if any states gather much of this important information. Not one state indicated it gathered or reported such contextual data. It is possible that the information is gathered elsewhere within state education departments, but it is likely that much of the desired information is not obtained or is not used in conjunction with assessment data.

Collecting contextual information is called for in the *Principles* because the information can be used in program evaluation, such as when interpreting achievement data. Additionally, while it would be inappropriate to justify low scores by reference to demographics, serious efforts at school reform require providing every student with an adequate and appropriate opportunity to learn. Thus, gathering contextual information is essential for using assessment results to improve programs rather than to simply report, praise or blame.

Standard 2: Assessments are fair.

Assessment systems must not limit students' present or future opportunities and must provide all students with a reasonable and fair opportunity to demonstrate their achievement. The *Principles* states: "Assessments are fair when every student has received equitable and adequate schooling, including culturally sensitive curriculum, instruction and assessment that encourage and support each student's learning....Assessment results accurately reflect a student's actual knowledge, understanding and achievement. Assessments are designed to minimize the impact of biases."

In some regards, states have made progress, particularly through bias and sensitivity review panels that often have the power to delete or revise items. Increasingly, states are aware of the need to provide adequate assessments to students with exceptional needs, but actual progress on such assessments has been limited. For students with Individual Education Plans (IEPs), this should soon change under the impetus of the recently revised Individuals with Disabilities Education Act (IDEA) federal legislation. The fairness standard also says that states do not make important decisions based on a single test score and that they provide students with opportunities to be assessed with multiple methods. On these issues, states are not making much progress.

2.1. States have implemented comprehensive bias review procedures. Bias in assessment renders an assessment invalid for the population against whom the assessment is biased. This is true not only because biased items fail to accurately measure all students' learning on that item, but also because biases can undermine how a student responds to an entire exam. Bias can include race, gender, socioeconomic class, culture, language, rural/urban, handicapping status, and sexual orientation. To guard against bias, committees -- with the authority to remove or modify items -- should examine individual items and the exam as a whole. Statistical procedures that can help detect biased items should also be used.

Most states have a bias review procedure. Bias reviews typically consider race and gender; some states reported considering disability or linguistic and cultural background; only a few states report considering other issues, such as socio-economic status.

Most have a separate bias review committee, though sometimes a content committee will examine items and the whole assessment for bias. For commercially published tests, states usually rely on bias review by the test maker, which often includes both committees (with unknown authority) and statistical studies. Thirteen of the states responding to the full FairTest survey reported doing statistical analyses of tests for bias, which should and sometimes does include studying tests both before and after administration.

In general, state and commercial exams appear to do fairly well in terms of identifying overtly biased items. Broader issues, such as the kinds of content in the composition of the test and the possible impact of the presence or absence of certain content (even if not overtly biased) on test takers, is studied in some states, but not in others (on this, we did not obtain much information).

2.2. Assessment results should be reported both for all students together and with disaggregated data for sub-populations. Failure to include all students in reports sends the message that they are less important and need not be considered. But it is also important to report disaggregated data in order to track the progress of groups which historically have not been well served by school systems.

A majority of states do some reporting of data disaggregated by demographic categories. States most commonly report by race and gender, while a few report socio-economic class. As noted below, states vary greatly in their reporting of students with IEPs or with limited English proficiency (LEP). In general, states need to do more to present disaggregated data, including at the district and school levels.

2.3. Adequate and appropriate accommodations and adaptations are provided for students with Individual Education Plans (IEP).

2.4. Adequate and appropriate accommodations and adaptations, including translations or developing assessments in languages other than English, are available for students with limited English proficiency (LEP).

States have only recently begun to consider including all students in their assessments. According to the National Center on Educational Outcomes (NCEO), many states still do not know how many students with IEPs are or are not assessed. Many states assess only a small percentage of their IEP students. The situation is often worse for students with LEP.

The 1997 reauthorization of the Individuals with Disabilities in Education Act (IDEA) requires states to develop standards for students with special needs that are coordinated with any state standards for all children; and to include students with IEPs in their accountability

systems, including assessments, with appropriate accommodations and, if necessary, alternate assessments. They are to be reported both in general reports and disaggregated. The new legislation therefore will bring the states closer in line with this standard. It is less certain that similar progress will be made in assessing students with LEP, as they are not included in the legislation.

A critical issue will be whether the assessments will be appropriate for the students. Not all students can reasonably be assessed with regular assessments. Some students require accommodations to make the results fair and meaningful. Still others may require alternate assessments. However, whether, or the extent to which, accommodations may alter the meaning of the assessment is not fully understood, and research is being done on this issue. Nonetheless, fairness requires that students with an IEP or who are LEP be assessed in terms of state standards and with appropriate assessments. The results should be included in regular reports wherever possible, as well as reported separately, so the success of programs for students with special needs can be evaluated. Requiring all students to be assessed and included in regular reports can also lessen the tendency to place some students in special programs so that they will not be assessed, enabling school or district scores to appear higher.

While states show a great range on this category, in general they do not yet properly include and assess IEP and LEP students.

FairTest attempted to obtain data on the percentage of students in each state with an IEP or who are LEP. The intent was to compare this with the percentage tested with LEP/IEP. However, too few states reported the first part for us to know for most states what percentages of students with IEP or LEP are not assessed. According to a recent NCEO report, many states do not know how many students are excluded. However, from the data available, it appears that large numbers of IEP and LEP students are not included in assessments in most states.

The accommodations or modifications available also vary greatly. The fewest tend to be available on commercial NRTs. Alternative assessments, such as the portfolio option used for more severely disabled students in Kentucky, are also very rare. Kentucky is the only state to assess all students with IEPs; no state assesses all students with LEP.

Though always desirable, assessments in languages other than English are particularly to be expected in states with high proportions or numbers of LEP students. California, Texas, New York, Florida, Illinois, Arizona, New Mexico, New Jersey, Michigan and Massachusetts have more than 40,000 students with LEP, and Washington and Oklahoma have over 25,000 LEP students. (See reports from George Washington University Evaluation Assistance Center East.) Only a few of these states provide assessments in languages other than English.

States vary in their reporting procedures for students with LEP and IEP. Some include them in regular reports, some publish separate reports, some do both, and some do neither. FairTest supports the approach of inclusion in regular reports and disaggregated reporting.

Finally, students with special needs should be included in the population for whom assessments are designed and in the population on whom tests are tried out. A few states reported doing this, though this question was not specifically asked. Additionally, professionals with knowledge of disability and language issues should be involved in developing the assessments.

*2.5. **Multiple methods of assessment are provided to students to meet needs based on different learning styles and cultural backgrounds.*** Students have varying learning styles and ways of expressing their knowledge and abilities. Different cultures reinforce different ways of organizing and demonstrating knowledge. Assessment should respond to these issues, as is recognized also in the *Standards for Educational and Psychological Measurement.*

Only a handful of states reported that they considered different learning styles or cultural variations, usually states that had included constructed-response items. It is likely that large-scale assessments, particularly exams, can only address this issue in a limited fashion. Even if a variety of methods are used in one exam, students can still be penalized for not doing well in one format compared with others. However, having multiple methods on an assessment at least conveys the need to use different methods in the classroom and provides some opportunities for students to use different modes of presenting knowledge.

*2.6. **Students are provided an adequate opportunity to learn about the assessment.*** Knowing about the format as well as the content of an assessment can be important to doing well. Knowledge about test methods should not be a source of score differences on measures of achievement. Thus, all students should be equally well prepared to use any methods employed on a large-scale assessment, and states should ensure that students are informed and prepared.

Most states make an effort to provide information to students, but the extent and quality of the information appears to vary greatly. As new assessment methods come into use, it is particularly important for states to ensure that students understand how to respond to those methods. Though states with new methods often provide examples for teachers to use with students, it is not clear whether these efforts actually ensure equity in format preparation among students.

Note: It is important to have a strong representation in the assessment development process of people from minority groups which will be assessed. Preferably, they would be over-represented in committees that design assessments and write and evaluate items, so that they can attain a critical mass to influence test construction. The survey did not address this issue.

Standard 3: Professional development.

The *Principles* explains, "Assessment systems depend on educators who understand the full range of assessment purposes, use appropriately a variety of suitable methods, work collaboratively, and engage in ongoing professional development to improve their capability as assessors."

States should ensure that incoming teachers have been adequately prepared to assess their students and that currently practicing teachers are competent assessors. States should provide or ensure that districts provide continuing professional development to meet this goal. Professional development is often enhanced by teachers' participation in developing and scoring performance tasks, so states should consider this value when they consider whether to contract out scoring.

The states are generally quite weak in providing adequate professional development in all aspects of assessment to teachers.

3.1. States have requirements for beginning teachers and administrators to be knowledgeable about assessment, including appropriate classroom practices. Without such requirements, schools of education may not require such preparation, leaving incoming teachers unable to adequately assess their students.

Most states have no assessment knowledge requirements for incoming teachers, and in particular they have no requirements for them to become competent in performance and classroom assessment. Licensing exams may have a few questions about assessment, but this is not a sufficient basis for assuming competence.

3.2. States provide sufficient professional development in assessment, including in classroom assessment. The state should ensure that teachers receive sufficient professional development in assessment. This support should be extensive and systematic. If states delegate this to districts, they should facilitate districts' ability to provide necessary professional development.

While most states provide some sort of professional development, most of it is neither extensive nor systematic. Various studies have suggested that even the best states find their efforts insufficient to meet demand when major reforms in standards or assessments occur. Since strengthened classroom assessment capabilities and restructured large-scale assessments are called for in the *Principles*, states need to do a great deal more to provide professional development and the opportunity for professional collaboration.

3.3. States survey educators about their professional development needs in assessment and evaluate their competence in assessment. These are means to determine what professional development is most needed. The evaluations should be done on an occasional and sampling basis to determine whether the professional development has succeeded and teachers are able to use assessments to support and evaluate student learning.

States rarely ask educators what they need regarding professional development in assessment, nor do they evaluate teacher competence in assessment. A few states have started to address this gap by surveying at least a sample of teachers about their needs and their practices as part of the state assessment program.

3.4. Teachers and other educators are involved in designing, writing and scoring assessments. These all provide opportunities for professional development, especially if the work is on more complex performance tasks or portfolios.

States often involve some teachers in writing items, often multiple-choice, on state-made assessments, but scoring of writing samples and constructed response or performance tasks is often contracted out. It appears that few teachers are actually involved in writing a state's items, and often the writing is of multiple-choice items, which fails to provide substantial professional development for classroom assessment. A few states have made an effort to engage a wide range of teachers in writing performance tasks, and others have teachers involved in scoring.

Two cautions. First, good tasks and items are not easy to write, and learning to write them takes time. Therefore, rigorous quality review of items is necessary. Second, the time to do this work needs to be organized so as not to detract from teaching.

States often cite cost as the reason to contract out scoring. FairTest recommends that when costs are estimated, the value of professional development be factored in. It may well be that the narrowness of state writing samples, for example, renders them not good vehicles for professional development, whereas scoring portfolios and complex tasks has often been found to be a powerful form of teacher education. While we generally support having teachers involved in scoring at least the more extended constructed-response items, it may be that states find it more effective to use professional development funds in other ways.

Standard 4: Public education, reporting and parents' rights.

Parents and the public have the right to be informed about assessments and assessment results and to have access to all reports. Thus, reports at times will need to be prepared in languages other than English. When new assessments are introduced, extensive public education may be necessary. This is both fair to parents and likely to be vital to the success of new assessments. It is useful for states to find out what parents and the public most want to know and to make sure that reports are understood by their intended audiences.

Parents and students also should have the right to review assessments and challenge scores or items they believe to be flawed. *A cult of secrecy surrounds testing which serves to conceal its limitations from public understanding and mystifies students as to what high quality work looks like and what is wanted on tests.* Some states are making progress toward openness, but much more needs to be done. Openness is worth the cost of writing more items.

4.1. Parents and community members are educated about the kinds of assessments used and the meaning and interpretation of assessment results. Parents and the public deserve to know what kinds of assessments are used and why, and to have results of assessments reported in a clear and comprehensible manner. This includes how to interpret the results and important inferences that can be drawn from them.

States typically provide public reports, and many provide guidance on using the results, but few states appear to make an extensive education effort about assessment beyond publishing test scores. States introducing new assessments usually do try to inform the public about them. Some states release items or provide examples of items and student work. In reporting assessment results, states should also provide contextual information about the schooling students received, though as noted earlier no states said they did this. States also should clearly state the limits of the data and cautions about common misuses and misinterpretations.

4.2. The state surveys parents/public to determine information they want on assessments and whether assessment reports are understandable. Reports should include information that parents and the public want, and reports should be understood by audiences. This requires public opinion research.

Fourteen of the states responding to the FairTest survey reported surveying as to what information the public wants. Of those 14, six also surveyed as to whether the reports are understandable.

4.3. Reports should be available in languages other than English if a sizeable number or significant percentage of the student population come from homes where another language is commonly used. Spanish-language reports would be the most common.

Only five states reported that they reported in languages other than English. Many states with large numbers of LEP students did not provide such reports.

4.4. Parents and/or students have the right to examine assessments, appeal assessment scores, or challenge flawed items. Parental review encourages openness. States should release items or tasks on a regular basis. Because scoring can be incorrect and items may be flawed, clear processes for appeals and challenges are necessary.

Most states allow parents to examine tests, often under secure conditions, and a few release all or many items for public review after each administration. Review of commercial NRTs is more limited and difficult, but is allowed in some states, indicating that contractual problems with the testmaker (a reason some states cited for not allowing test review) can be resolved.

Eleven states reported on the FairTest survey that they allow item challenges or score appeals. Score appeals are more likely to be allowed on writing samples and constructed-response items, which are scored by people rather than machines, and on high school exit exams, where mistakes have more serious consequences.

Note: For a variety of reasons, some parents object to all or some kinds of large-scale testing. *Ten states reported allowing parents to exclude their children from an exam.* Some said requests for exemptions were growing, though the number remained small. A few even

included the high school exit exam in the tests covered by such exemption policies, but in some of these the state said it would ask the parent to sign a form indicating awareness that the child would not receive a standard diploma if she or he did not take and pass the test. In such cases, given the relatively older age of the children and the consequences, it is probably wise for the child to also assent to opting out.

This was not an issue raised in the *Principles*. In the face of tests that may be more harmful than helpful, a parental right to exempt children may be reasonable. A caution should be raised, however, that schools do not use such a right as a lever to persuade parents of low-scoring children to opt out -- that is, to push them out.

Standard 5: System review and improvement.

States should regularly review their assessment programs in order to assure the quality of the system, to prevent or remedy harmful consequences of test use, to support beneficial consequences, and to provide information useful for improving the system. A comprehensive review would include the factors discussed in the *Principles*. This would include the quality and effectiveness of bias reduction, the extent of inclusion, professional competence in assessment, and the quality of public reporting. Including assessment as part of a review of a state's entire educational program probably makes more sense than just conducting separate reviews of assessment.

While most states conduct some form of review, their review practices are limited and important areas are often not addressed.

A comprehensive review of an assessment used for public information or accountability would help determine if:

- the data are accurate;
- the accountability system is relevant to important issues and actually reports what it says it reports (e.g., a report on writing is based on educationally valid understandings of writing);
- any impact the assessment has is at least neutral, preferably positive, and certainly not harmful to curriculum, instruction, student progress, or the cognitive and emotional development of children; and
- assessments measure in a balanced manner all important aspects of the standards or curriculum on which they are based and thus assess critical thinking and cognitively complex activity within and across subject areas.

Few states can provide data about their assessment program with respect to these key issues. In an era in which testing is proposed as a fundamental tool for school reform, states often can report little more than that scores are increasing or decreasing. They often cannot even be sure whether increasing scores are based on real learning gains or teaching to the test. Additionally, though most states have powerful leverage over district practices, such as through state constitutions, few states have evaluated their districts' assessment practices.

There is a further issue: the values and assumptions that underlie state reviews. For example, some states have concluded that the multiple-choice tests they use are appropriate for young children, contrary to the professional consensus in the field. Others claim that their multiple-choice tests can assess complex and critical thinking, which suggests that they and their critics may hold different conceptions of critical thinking.

We were able to examine a few independent and self-evaluations of states. The conceptual structures and values of the evaluators are clearly important in how they frame their approaches. Acceptance of traditional psychometric values and concepts, which underlie traditional exams, produce different evaluative conclusions than those based on different views of learning (such as constructivist or social constructivist models) or of the goals of schooling. *Reviewers need to make explicit and defend the perspectives, assumptions and values which undergird their reviews.*

Improving the evaluation process should be a priority in most states. The reviews must seriously and critically engage the underlying concepts of the state assessment programs.

5.1. The assessment system is regularly reviewed.

Twenty-eight of the forty-three states which responded to the FairTest survey reported that they have some sort of review process. All states should have comprehensive review procedures.

5.2. The review includes participation by various stakeholders and evaluation by independent experts. Participation by the public and independent experts helps ensure credibility and brings diverse views to the review process. While test developers or contractors should participate in evaluating the system, they are not independent evaluators.

Twenty-three states reported involvement by educators, 10 by one or more community sectors, 16 by SEA staff, three by test contractors. Three employed independent, outside experts. In general, the range of stakeholders involved is limited, and few states arrange for outside evaluation with any regularity, if at all. A few states have studied their systems in great detail and used outside experts as well as at least some stakeholders. These states are often those which have begun to develop fundamentally new assessment systems, such as Kentucky and Vermont.

5.3. The review studies how well the system actually is aligned to standards.

While some states reported studies as to the match between state standards or curriculum and the assessments, the reviews often fail to evaluate how well the assessment measures all aspects of the standards. In most cases, the studies appear to focus on whether test content is included in the standards; this is particularly the case when the match is to a commercially published test.

5.4. The review studies the impact of the assessment(s) on curriculum and instruction. Assessments can have a variety of consequences for school practice and the actual curriculum and instruction students receive. These consequences -- desired and undesired, beneficial and harmful -- should be studied in order to eliminate problems and enhance strengths.

Only 13 states reported studying the impact of state-mandated assessments on curriculum and instruction. Some states reported increased scores on the assessments as a positive impact. While teaching to the test can be positive if it does not narrow instruction in harmful ways, without further study states cannot be sure how much gain is real learning and how much is test-score inflation on a too-narrow test that is taught to in too-narrow ways.

5.5. The review studies whether assessments assess critical thinking or the ability to engage in cognitively complex work within a subject.

A mere five states reported studying whether the assessments measured critical thinking or cognitive complexity. Most state assessments are dominated by methods known to have limited capacity to assess critical thinking, but most states do not investigate this issue.

5.6. Reviews for assessments at grade 3 or below study whether the assessments are developmentally appropriate. Experts on the education of young children have advocated that assessment be "developmentally appropriate," that is, reasonable for the range of capabilities and ways of learning of students through age 8 (see Bredekamp).

Most states which test at or below grade 3 claim to have studied the assessments for developmental appropriateness, but it appears some of these studies may not include critical issues raised by experts on this age group. Of 24 states with mandated assessments at grade 3 or earlier, two reported studying them for developmental appropriateness (the actual number may be slightly higher, as not all states responded to the full FairTest survey). Guidelines for developmentally appropriate assessment for young children have cautioned against the use of multiple-choice tests, but some states have said they have reviewed their multiple-choice tests for appropriateness. It would appear, therefore, that those guidelines have not been used in selecting or evaluating the assessments.

5.7. Reviews study the impact of assessment programs on student progress and particularly the impact of any high-stakes tests, such as high school exit exams, on graduation rates. If graduation tests, for example, reduce the graduation rate or do so differently for different population groups, the state should know this and take appropriate steps to address the problem.

Seventeen states have mandatory high school exit exams. Of these, 12 responded to the FairTest survey and only four of them reported studying the impact on high school graduation. Since the use of single exams as a hurdle to high school graduation or grade promotion violates professional standards, states that persist in doing so should study the consequences of those exams. Preferably, the studies should be done by independent

contractors not invested in the outcomes of such studies.

5.8. Reviews study the technical quality of assessments. Technical considerations, most importantly validity, but also generalizability, reliability, bias, and scoring procedures, should always be studied. Validity is fundamental, and overlaps with the topics addressed above, including the match with standards, assessment of critical thinking, impact on curriculum and instruction and on high school graduation rates, and bias. Gathering evidence about the validity of an assessment is a continuing process rather than a one-time effort.

Far too few states conduct technical studies of their assessments. Fourteen states reported doing technical studies. Technical studies on commercial tests are usually done by the publishers. Technical and consequential aspects of validity are complementary and both must be studied. This survey did not investigate the nature of the technical studies to determine what elements were included in the studies, nor was the quality of the studies evaluated.

5.9. The state reviews local assessment practices. This should include use of surveys regarding classroom, school or district assessment practices. This standard suggests that states have a responsibility to oversee district assessment practices in order to help prevent harmful practices and to support improvement.

Very few states survey to find out about district, school or teacher assessment practices, or review or evaluate local assessment practices. Four reported that they review district assessments, and one reported reviewing school assessments.

5.10. Reviews help guide improvements in the assessment system that will bring the program more in line with the **Principles and Indicators.** Studies of the system should provide information useful for improving the system. The *Principles and Indicators* should be used to help shape the changes in a beneficial direction.

Few states that are revising their assessment systems reported using studies of the current or previous system in making revisions. Some state changes represent progress toward the Principles. *Others do not or are even steps backwards.*

B. Standards for Evaluating State Assessment Systems

Standard 1: Assessment supports important student learning.

1.1. Assessments are based on and aligned with standards.

1.2. Multiple-choice and very-short-answer (e.g., "gridded-in") items are a limited part of the assessments; and assessments employ multiple methods, including those that allow students to demonstrate understanding by applying knowledge and constructing responses.

1.3. Assessments designed to rank order, such as norm-referenced tests (NRT), are not used or are not a significant part of the assessment system.

1.4. The test burden is not too heavy in any one grade or across the system.

1.5. High stakes decisions, such as high school graduation for students or probation for schools, are not made on the basis of any single assessment.

1.6. Sampling is employed to gather program information.

1.7. The evaluation of work done over time, e.g., portfolios, is a major component of accountability and public reporting data.

1.8. Students are provided an opportunity to comment on or evaluate the instruction they receive and their own learning.

1.9. Appropriate contextual information is gathered and reported with assessment data.

Standard 2: Assessments are fair.

2.1. States have implemented comprehensive bias review procedures.

2.2. Assessment results should be reported both for all students together and with disaggregated data for sub-populations.

2.3. Adequate and appropriate accommodations and adaptations are provided for students with Individual Education Plans (IEP).

2.4. Adequate and appropriate accommodations and adaptations, including translations or developing assessments in languages other than English, are available for students with limited English proficiency (LEP).

2.5. Multiple methods of assessment are provided to students to meet needs based on different learning styles and cultural backgrounds.

2.6. Students are provided an adequate opportunity to learn about the assessment.

Standard 3: Professional development.

3.1. States have requirements for beginning teachers and administrators to be knowledgeable about assessment, including appropriate classroom practices.

3.2. States provide sufficient professional development in assessment, including in classroom assessment.

3.3. States survey educators about their professional development needs in assessment and evaluate their competence in assessment.

3.4. Teachers and other educators are involved in designing, writing and scoring assessments.

Standard 4: Public education, reporting, and parents' rights.

4.1. Parents and community members are educated about the kinds of assessments used and the meaning and interpretation of assessment results.

4.2. The state surveys parents/public to determine information they want on assessments and whether assessment reports are understandable.

4.3. Reports should be available in languages other than English if a sizeable number or significant percentage of the student population come from homes where another language is commonly used.

4.5. Parents and/or students have the right to examine assessments, appeal assessment scores, or challenge flawed items.

Standard 5: System review and improvement.

5.1. The assessment system is regularly reviewed.

5.2. The review includes participation by various stakeholders and evaluation by independent experts.

5.3. The review studies how well the system actually is aligned to standards.

5.4. The review studies the impact of the assessment(s) on curriculum and instruction.

5.5. The review studies whether assessments assess critical thinking or the ability to engage in cognitively complex work within a subject.

5.6. Reviews for assessments at grade 3 or below study whether the assessments are developmentally appropriate.

5.7. Reviews study the impact of assessment programs on student progress and particularly the impact of any high stakes tests, such as high school exit exams, on graduation rates.

5.8. Reviews study the technical quality of assessments.

5.9. The state reviews local assessment practices.

5.10. Reviews help guide improvements in the assessment system that will bring the program more in line with the *Principles and Indicators*.

C. Scoring Guide

The FairTest evaluation focuses on the primary characteristics described below. States' scores are based primarily on their current programs, but on occasion changes that are currently being implemented were considered.

Level 1. State assessment system needs a complete overhaul. Such a state system exhibits *three or more* of the following negative characteristics:

- Uses all or almost all multiple-choice testing;
- Tests all students in one or more grades with a norm-referenced test;
- Has a single exam as a high school exit or grade-promotion requirement; or
- Exhibits generally poor performance on the other standards.

Level 2. State assessment system needs many major improvements. Such a state system has *two* of the following negative characteristics:

- Uses all or almost all multiple-choice testing;
- Tests all students in one or more grades with a norm-referenced test;
- Has a single exam as a high school exit or grade-promotion requirement; or
- Exhibits generally poor performance on the other standards.

Level 3. State assessment system needs some significant improvements. Such a state system has some positive attributes but still has *one* of the following negative characteristics:

- Uses all or almost all multiple-choice testing;
- Tests all students in one or more grades with a norm-referenced test;
- Has a single exam as a high school exit or grade-promotion requirement; or
- Exhibits generally poor performance on the other standards.

Level 4. State assessment system needs modest improvement. Such a state system generally performs well across the standards, has none of the major problems described at previous levels, but does not show all the characteristics of a model system, including use of sampling and classroom-based assessments for accountability and public reporting.

Level 5. A model system. Such a state system performs well across all the standards, including use of sampling and classroom-based assessments as significant portions of accountability and public reporting. It may need minor improvements in some areas.

Not scorable. The state does not have an assessment system and does not mandate any assessments for districts to use, or is otherwise not scorable.

Discussion. This scoring guide gives the most weight to Standard 1. If an assessment system does not support high quality teaching and learning, it should be completely overhauled. The presence of some ameliorating characteristics such as limited use of NRT (e.g., only one grade and subject) or alternatives to the graduation requirement, or some other significant positive attributes from the other standards can move a state up a level.

D. STATE DATA TABLE

1996-97

STATE	level	m-c	nrt	grad test	writing	purposes
AL	1	1	1	1	1	1,4,6
AK	1	1	1	3	1	1,2,6
AZ	1	1	1			1,2,6
AR	2	1	1	3		1,2,6
CA*	2	2	1			1,5,6
CO	4	2			1,3	1,6
CT	4	2/3			1,3	1,2,6
DE**	0	4**		3,2	1	1
FL	1	1	1	1	1	1,2,4
GA	1	1	1	1	1	1,2,3,6
HI	1	1	1	1		1,2,6
ID	2	2	1		1	1,2
IL	3	1	1		1	1,4
IN	2/1	2	1		1	1,2,3,4,6
IA	0					
KS	3/4	2			1	1,2
KY	4/3	3	1		2	1,2,3,4
LA	1	2	1	1	1	1,2,5
ME	4	4			1	1,2,5
MD	3	3	2	1	1	1,2,3,4,6
MA	2	1	1	3		2
MI	3	2		4	4	1,2,3,4,5,6
MN	2	1		2	1	1
MS	1	1	1	1	1	1,2,4,6
MO^	4/3^	1			1,3	1,2,4,6

STATE	level	m-c	nrt	grad test	writing	purposes
MT	2	1	1			1,2
NE	2	1	1			2
NE	2	1	1	1	1	1,2,6
NH	4	2			1	1,2
NJ	2	2		1	1	1,2,4,5,6
NM	1	2	1	1	1,2	1,2,6
NY	2	2		1	1	1,2,3,4,5,6
NC	1	2	2	1	1	1,2,3,4,6
ND	2	1	1			1,2,4,6
OH	2	1		1	1	1,2,6
OK	2/1	1	1		1	1,2,4,5
OR	3	2			3	1,2,6
PA	3	2			1	1,2,3
RI	3	2/3	1		1	1,2,6
SC	1	1	1	1	1	1,2,3,4,5,6
SD	2	1	1			1,2
TN	1	1	1	1	1	1,2,3,4,6
TX	2	1		1	1	1,2,3,4,5,6
UT	1	1	1			1,2,5,6
VT	5	4			2	1,2
VA	1	1	1	1	1	1,2,5,6
WA	2	2	1		1	1,2,3
WV	1	1	1	4	1	1,2,4,5,6
WI	2	2	1		1	1,2,4,6
WY+	0	4				1

Coding and notes follow on next two pages.

Coding of table

level = the level of the state program according to the FairTest scoring guide
- 1 = needs a complete overhaul
- 2 = needs many major improvements
- 3 = needs some significant improvements
- 4 = needs modest improvement
- 5 = model system
- 0 = no state system and no state mandate for particular district testing; or otherwise not scorable

mc = multiple-choice, excluding writing assessment
- 1 = all/almost all m-c
- 2 = majority m-c
- 3 = minority m-c
- 4 = no/almost no m-c

nrt = use of a norm-referenced test (NRT)
- 1 = uses an NRT
- 2 = uses an NRT, but on a sampling basis

grad test = graduation test
- 1 = has a test and passing it is required for graduation
- 2 = has a required graduation test, but also an acceptable alternative
- 3 = state plans to require a graduation test but does not now have one
- 4 = has a graduation test, but passage is not required for diploma

writing = states have a writing assessment
- 1 = write to a prompt
- 2 = portfolio
- 3 = multiple choice
- 4 = anything else for writing

purposes = purposes for the test
- 1 = improve curriculum and instruction
- 2 = program evaluation/public reporting
- 3 = rewards for schools/districts
- 4 = sanctions for schools/districts
- 5 = rewards or sanctions for students other than high school graduation
- 6 = student diagnosis

Notes:

Data is from 1996-97 school year, except 1995-96 for Arkansas, Connecticut, Florida, Maryland, Mississippi, Ohio, which did not respond to FairTest survey.

In the "level" column, use of a slash (/), as in 4/3, indicates that the system is on the border; the first number is the direction in which the state appears to be leaning. In this column, numbers separated by a comma indicate a system whose parts (current, or current and being implemented) require separate evaluation.

In the multiple-choice ("m-c") column, use of a slash (/) indicates we could not precisely determine the proportions of multiple-choice items used on state assessments.

* California pays districts to test voluntarily, mostly with NRTs (hence a 2) and has other exams that are criterion-referenced with some constructed-response (hence a 3).

** Delaware assessed only writing 1996-97, not a full state testing program, hence a 0. Its new program is still being designed, but it will include norm-referenced tests and a high school exit exam (which will allow for alternatives) hence a 2.

^ Missouri's incoming program appears likely to score at a level 4; the current program, which relies primarily on criterion-referenced multiple-choice items but employs sampling, rates a 3.

+ Wyoming assessed only employment readiness in 1996-97, and that on a sampling basis, making it really a state without a state assessment system.

Methodology

A) Sources of information

FairTest began with the 1994-95 CCSSO/NCREL survey, published in May 1996. We matched the data available from that survey to the *Principles* and discovered that many areas of the *Principles* were not covered by that survey.

We then analyzed the *Principles* to extract indicators relevant to large-scale assessments or state-level practices. We excluded areas in which information was not likely to be available. From the remainder, we constructed a fairly long survey. We asked two state assessment directors to look over the survey. In addition to suggested clarifications, one advised us that the survey was too long. While we condensed it somewhat, we decided to attempt to gather all the information we could. We mailed the survey to all 50 states in the summer of 1996. (Washington, DC, is not included in the CCSSO/NCREL survey; we sent DC both that survey and ours, but they did not respond, so they are not included in the report.) A copy of the final survey sent to all 50 states is in Appendix D.

Responses began to come in, but a few states indicated they would not participate. In the fall of 1996, we sent a follow-up letter. In early 1997, we checked with a number of states which had not replied to determine whether they would be amenable to responding to a shortened version of the survey, and a number indicated yes. The cuts were made in areas in which we had not received much information in the surveys that had been returned or in areas we decided were of less importance. A few states answered the short form questions over the telephone, rather than respond on paper. (A copy of the short-form survey is in Appendix E.) As a result of the change in the form, because some items were left blank by states, and because some states did not respond to the survey at all, the extent of the information varies from state to state.

FairTest also relied on other sources of information. We used AFT and the CCSSO reports to summarize whether a state had standards and in what subjects. News reports in media such as *Education Week* alerted us to possible changes in state assessments that we then checked, sometimes by telephone. For each state report, we list the data sources used.

Based on completed surveys, we wrote draft descriptive summaries of each state. We sent these to states to have them checked for accuracy. In a few cases, either many significant changes in the state program had occurred since the survey was first filled out or the state suggested many changes in the description. In those cases we redrafted and sent the survey back to the state for further review. In a few cases, information on standards was added after the state had checked off the descriptive draft.

For states that did not respond to the survey, we relied solely on other sources, primarily the NCREL/CCSSO survey for 1995-96 (released in June 1997), plus the AFT and CCSSO reports on standards. As a result, significant areas are not discussed for those states.

Despite our efforts to collect data on all aspects of the *Principles* and to verify that data, we recognize a series of potential problems:

- Variability in the thoroughness of state responses.
- Some information was not rechecked with the state.
- The information received depends in part on the person sending it. Occasionally

we were told that material we found in other reports had never been true. Such problems may affect this report as well, though we, like others, have attempted to confirm information.

There are state assessments that are not included in this survey. For example, some states require particular tests to be used for entrance into and exit out of programs for LEP students, but no state reported those assessments as part of the state testing program. There also may be other mandates to districts that states do not report.

Despite these potential problems, we are very confident that the data are substantially accurate and that having additional or in some cases more recent data would not alter the national findings in any significant way and only rarely would affect a state report.

Having obtained and checked the data, we subjected it to an evaluation based on the *Principles*. The grounds for evaluation and a rubric for ratings are discussed in the first parts of the section on state findings. Thus, the evaluations are FairTest's and not those of the National Forum on Assessment, which wrote the *Principles*.

B) Implications for future surveys and studies.

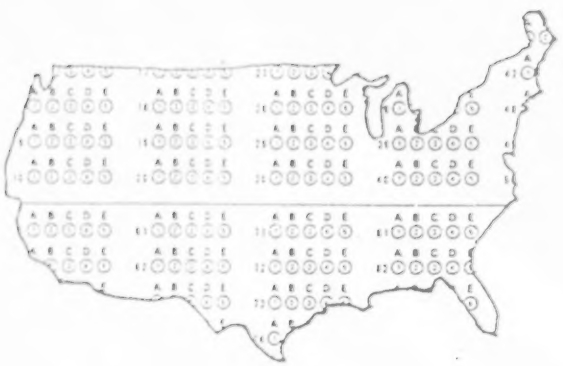
While the CCSSO/NCREL survey is a valuable source of information, the FairTest report includes many important areas that have not been studied by the CCSSO/NCREL survey. Topics central to the *Principles*, such as program review and evaluation, bias reduction, and professional development, are often either not included or included in only a very cursory fashion. It also is difficult to disentangle some CCSSO/NCREL data. For example, states often included their writing samples in response to questions about whether they have non-multiple-choice items in their assessments, making it difficult to determine if they had any other form of constructed-response or performance items. FairTest hopes that future CCSSO/NCREL surveys will include questions asked in the FairTest survey, making it an even more comprehensive source of data.

A major limitation of the FairTest and other surveys is the ability to use data to evaluate the actual quality of state assessments; standards; bias reduction, equity and professional development efforts; public reporting; and reviews. This is not a limitation that can readily be solved through survey methodology. Rather, it requires a more detailed qualitative analysis of state assessment programs. There does not appear to be a truly independent and representative body to undertake that important work.

FairTest's evaluations and conclusions are based on applying findings from a range of research on assessment to the available data from the states. For example, if state A uses a high-stakes, mostly multiple-choice test, FairTest's critique is based on research about high-stakes testing and multiple-choice tests and their educational impact. It is not based on a specific study of the consequences in state A. Such studies are needed, but as the FairTest survey shows, few states conduct them.

TESTING OUR CHILDREN

A Report Card on State Assessment Systems



by Monty Neill and the Staff of FairTest

FairTest: The National Center for Fair & Open Testing

Testing Our Children: A Report Card on State Assessment Systems $_____
(FairTest, 1997: 250 pp.) Individual copies for $30; 5 for $125.

Testing Our Children: Executive Summary $_____
(FairTest, 1997: 40 pp.) Individual copies for $10; 5 for $40.

Principles and Indicators for Student Assessment Systems $_____
Individual copies for $10; 10 for $80; 50 for $350; 100 for $600.

FairTest Examiner, published quarterly; $45/year (institutions), $30/year (individuals) $_____

Name _____ TOTAL $_____

Address _____

City _____ State _____ ZIP _____

42

Make checks payable to: FairTest, 342 Broadway, Cambridge, MA 02139

- ❒ **FAIRTEST EXAMINER**, published quarterly ... $30/yr (individuals), $45/yr (institutions). Outside U.S. add $5 postage. Back issues $7: Vol. ___ No. ___ $ _____
- ❒ ***THE SAT COACHING COVER-UP**, *by Sarah Stockwell, Bob Schaeffer and Jeff Lowenstein* (1991: 48 pp.) ... $9.95 $ _____
- ❒ **STANDING UP TO THE SAT**, *by John Weiss, Barbara Beckwith and Bob Schaeffer* (1989: 192 pp.) ... $6.95 $ _____
- ❒ ***SAT FACT PACKET**, *fact sheets & articles* (1994) ... $5.00 $ _____
- ❒ **THE CASE AGAINST THE SAT**, *by James Crouse and Dale Trusheim* (U. of Chicago Press, 1988: 224 pp.) ... $22.95 $ _____
- ❒ ***FALLOUT FROM THE TESTING EXPLOSION**, *by Noe Medina and Monty Neill*, 3rd ed. (1990: 77 pp.) ... $11.95 $ _____
- ❒ **AUTHENTIC ASSESSMENT IN ACTION**, *by Linda Darling-Hammond, J. Ancess and B. Falk* (Teachers College Press, 1995: 304 pp.) ... $24.95 $ _____
- ❒ **THE REIGN OF ETS**, *by Allan Nairn and Ralph Nader* (CSRL, 1980: 550 pp.) ... $30.00 $ _____
- ❒ **FAIR EMPLOYMENT STRATEGIES IN HUMAN RESOURCE MANAGEMENT**, *ed. Richard S. Barrett* (Quorum Books, 1996: 336 pp.) ... $59.95 $ _____
- ❒ **THE PRIMARY LEARNING RECORD**, *by Hilary Hester, et al.* (CLPE, 1993: 73 pp.) ... $35.00 $ _____
- ❒ ***STANDARDIZED TESTS AND OUR CHILDREN:** A Guide to Testing Reform, *by FairTest* (1991: 32 pp.) ... $4; 5 for $15; 10 for $20; 50 for $50; 100 for $75. English, Spanish and New York editions available. $ _____
- ❒ **THE MISMEASURE OF MAN**, *by Stephen Jay Gould* (W.W. Norton, 1991: 352 pp.) ... $9.95 $ _____
- ❒ ***BEYOND STANDARDIZED TESTS: ADMISSIONS ALTERNATIVES THAT WORK**, *by Amy Allina with FairTest staff* (1987: 18 pp.) ... $5.50 $ _____
- ❒ **TESTING AFRICAN AMERICAN STUDENTS**, *ed. Asa Hilliard* (Third World Press, 1991: 185 pp.) ... $12.95 $ _____
- ❒ ***PRINCIPLES AND INDICATORS FOR STUDENT ASSESSMENT SYSTEMS**, *by the National Forum on Assessment* (FairTest, 1995: 36 pp.) ... $10; 10 for $80; 50 for $350; 100 for $600. $ _____
- ❒ **TESTING FOR LEARNING:** How New Approaches to Evaluation Can Improve American Schools, *by Ruth Mitchell* (Free Press, 1992: 222 pp.) ... $22.95 $ _____
- ❒ ***IMPLEMENTING PERFORMANCE ASSESSMENTS:** A Guide to Classroom, School and System Reform, *by FairTest* (1995: 56 pp.) ... $6; 5 for $25; 10 for $40; 25 for $75; 50 for $125; 100 for $175. $ _____
- ❒ ***ANNOTATED BIBLIOGRAPHIES**, *by FairTest* $ _____
  Bilingual Assessment (1995) ... $5.00 # _____
  Performance Assessment (1996) ... $5.00 # _____
  Assessment of Young Children (1993) ... $3.00 # _____
  SAT: Bias and Misuse (1995) ... $8.00 # _____
- NEW! ***TESTING OUR CHILDREN:** A Report Card on State Assessment Systems, *by Monty Neill and the staff of FairTest* (FairTest, 1997: 250 pp.) ... $30.00
  Executive Summary (40 pp.) ... $10.00 $ _____

CONTRIBUTIONS

I agree there's no better way to be part of the testing reform movement than supporting FairTest! Enclosed is my tax-deductible contribution.#

- ❒ **FAIRTEST ASSOCIATE**
  $50—includes *Examiner* and a free copy of one of the starred (•) FairTest publications.
- ❒ **FAIRTEST SPONSOR**
  $75—includes *Examiner* subscription and two starred (•) FairTest publications.
- ❒ **FAIRTEST SUSTAINER**
  $100 or more—includes above, plus a 20% discount on all other publications.
- ❒ **OTHER**
  $_____

TOTALS

PUBLICATIONS TOTAL
$ _____

CONTRIBUTIONS TOTAL
$ _____

CHECK TOTAL
$ _____

Shipping and handling included.

Make payable to FairTest and mail to 342 Broadway, Cambridge, MA 02139.

#All contributions to FairTest are fully tax-deductible, except for the value of complimentary books or reports.

NAME _____

ADDRESS _____

CITY _____ STATE _____ ZIP _____